

0590  
1026#2  
OIEP

## RAW SEQUENCE LISTING

DATE: 12/04/2001

PATENT APPLICATION: US/09/973,025

TIME: 11:35:30

Input Set : N:\Crf3\RULE60\09973025.txt

Output Set: N:\CRF3\12042001\I973025.raw

## SEQUENCE LISTING

## 3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: MAERTENS, GEERT

6 BOSMAN, FONS

7 DE MARTYNOFF, GUY

8 BUYSE, MARIE-ANGE

10 (ii) TITLE OF INVENTION: PURIFIED HEPATITIS C VIRUS ENVELOPE

11 PROTEINS FOR DIAGNOSTIC AND THERAPEUTIC USE

13 (iii) NUMBER OF SEQUENCES: 111

15 (iv) CORRESPONDENCE ADDRESS:

16 (A) ADDRESSEE: NIXON &amp; VANDERHYE P.C.

17 (B) STREET: 1100 NORTH GLEBE ROAD

18 (C) CITY: ARLINGTON

19 (D) STATE: VIRGINIA

20 (E) COUNTRY: U.S.A.

21 (F) ZIP: 22201-4714

23 (v) COMPUTER READABLE FORM:

24 (A) MEDIUM TYPE: Floppy disk

25 (B) COMPUTER: IBM PC compatible

26 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

27 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

29 (vi) CURRENT APPLICATION DATA:

C--&gt; 30 (A) APPLICATION NUMBER: US/09/973,025

C--&gt; 31 (B) FILING DATE: 10-Oct-2001

32 (C) CLASSIFICATION: 435

34 (vii) PRIOR APPLICATION DATA:

35 (A) APPLICATION NUMBER: 08/612,973

36 (B) FILING DATE: 1996-03-11

37 (viii) ATTORNEY/AGENT INFORMATION:

38 (A) NAME: BYRNE, THOMAS E.

39 (B) REGISTRATION NUMBER: 32,205

40 (C) REFERENCE/DOCKET NUMBER: 1487-10

42 (ix) TELECOMMUNICATION INFORMATION:

43 (A) TELEPHONE: (703) 816-4000

44 (B) TELEFAX: (703) 816-4100

48 (2) INFORMATION FOR SEQ ID NO: 1:

50 (i) SEQUENCE CHARACTERISTICS:

51 (A) LENGTH: 21 base pairs

52 (B) TYPE: nucleic acid

53 (C) STRANDEDNESS: single

54 (D) TOPOLOGY: linear

56 (ii) MOLECULE TYPE: cDNA

58 (iii) HYPOTHETICAL: NO

C--&gt; 60 (iv) ANTI-SENSE: NO

65 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

67 GGCATGCAAG CTTAATTAAT T

69 (2) INFORMATION FOR SEQ ID NO: 2:

ENTERED

21

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/973,025

DATE: 12/04/2001

TIME: 11:35:30

Input Set : N:\Crf3\RULE60\09973025.txt

Output Set: N:\CRF3\12042001\I973025.raw

```

71      (i) SEQUENCE CHARACTERISTICS:
72          (A) LENGTH: 68 base pairs
73          (B) TYPE: nucleic acid
74          (C) STRANDEDNESS: single
75          (D) TOPOLOGY: linear
77      (ii) MOLECULE TYPE: cDNA
79      (iii) HYPOTHETICAL: NO
C--> 81      (iv) ANTI-SENSE: NO
85      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
87 CCGGGGAGGC CTGCACGTGA TCGAGGGCAG ACACCATCAC CACCATCACT AATAGTTAAT      60
89 TAACTGCA                                                                    68
91 (2) INFORMATION FOR SEQ ID NO: 3:
93      (i) SEQUENCE CHARACTERISTICS:
94          (A) LENGTH: 642 base pairs
95          (B) TYPE: nucleic acid
96          (C) STRANDEDNESS: single
97          (D) TOPOLOGY: linear
99      (ii) MOLECULE TYPE: cDNA
101     (iii) HYPOTHETICAL: NO
C--> 103     (iv) ANTI-SENSE: NO
106     (ix) FEATURE:
107         (A) NAME/KEY: CDS
108         (B) LOCATION: 1..639
110     (ix) FEATURE:
111         (A) NAME/KEY: mat_peptide
112         (B) LOCATION: 1..636
115     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
117 ATG CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT TTA CTG TCC TGT      48
118 Met Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys
119 1          5          10          15
121 CTG ACC ATT CCA GCT TCC GCT TAT GAG GTG CGC AAC GTG TCC GGG ATG      96
122 Leu Thr Ile Pro Ala Ser Ala Tyr Glu Val Arg Asn Val Ser Gly Met
123          20          25          30
125 TAC CAT GTC ACG AAC GAC TGC TCC AAC TCA AGC ATT GTG TAT GAG GCA      144
126 Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala
127          35          40          45
129 GCG GAC ATG ATC ATG CAC ACC CCC GGG TGC GTG CCC TGC GTT CGG GAG      192
130 Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys Val Arg Glu
131          50          55          60
133 AAC AAC TCT TCC CGC TGC TGG GTA GCG CTC ACC CCC ACG CTC GCA GCT      240
134 Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala
135 65          70          75          80
137 AGG AAC GCC AGC GTC CCC ACC ACG ACA ATA CGA CGC CAC GTC GAT TTG      288
138 Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu
139          85          90          95
141 CTC GTT GGG GCG GCT GCT CTC TGT TCC GCT ATG TAC GTG GGG GAT CTC      336
142 Leu Val Gly Ala Ala Leu Cys Ser Ala Met Tyr Val Gly Asp Leu
143          100          105          110
145 TGC GGA TCT GTC TTC CTC GTC TCC CAG CTG TTC ACC ATC TCG CCT CGC      384

```

## RAW SEQUENCE LISTING

DATE: 12/04/2001

PATENT APPLICATION: US/09/973,025

TIME: 11:35:30

Input Set : N:\Crf3\RULE60\09973025.txt

Output Set: N:\CRF3\12042001\I973025.raw

```

146 Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile Ser Pro Arg
147      115      120      125
149 CGG CAT GAG ACG GTG CAG GAC TGC AAT TGC TCA ATC TAT CCC GGC CAC      432
150 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His
151      130      135      140
153 ATA ACA GGT CAC CGT ATG GCT TGG GAT ATG ATG ATG AAC TGG TCG CCT      480
154 Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro
155 145      150      155      160
157 ACA ACG GCC CTG GTG GTA TCG CAG CTG CTC CGG ATC CCA CAA GCT GTC      528
158 Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro Gln Ala Val
159      165      170      175
161 GTG GAC ATG GTG GCG GGG GCC CAT TGG GGA GTC CTG GCG GGC CTC GCC      576
162 Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala Gly Leu Ala
163      180      185      190
165 TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT TTG ATT GTG ATG CTA      624
166 Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile Val Met Leu
167      195      200      205
169 CTC TTT GCT CTC TAATAG      642
170 Leu Phe Ala Leu
171      210
174 (2) INFORMATION FOR SEQ ID NO: 4:
176      (i) SEQUENCE CHARACTERISTICS:
177          (A) LENGTH: 212 amino acids
178          (B) TYPE: amino acid
179          (D) TOPOLOGY: linear
181      (ii) MOLECULE TYPE: protein
183      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
185 Met Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys
186 1      5      10      15
188 Leu Thr Ile Pro Ala Ser Ala Tyr Glu Val Arg Asn Val Ser Gly Met
189      20      25      30
191 Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val Tyr Glu Ala
192      35      40      45
194 Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys Val Arg Glu
195      50      55      60
197 Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr Leu Ala Ala
198 65      70      75      80
200 Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His Val Asp Leu
201      85      90      95
203 Leu Val Gly Ala Ala Ala Leu Cys Ser Ala Met Tyr Val Gly Asp Leu
204      100      105      110
206 Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile Ser Pro Arg
207      115      120      125
209 Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr Pro Gly His
210      130      135      140
212 Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro
213 145      150      155      160
215 Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro Gln Ala Val
216      165      170      175

```

## RAW SEQUENCE LISTING

DATE: 12/04/2001

PATENT APPLICATION: US/09/973,025

TIME: 11:35:30

Input Set : N:\Crf3\RULE60\09973025.txt

Output Set: N:\CRF3\12042001\I973025.raw

218 Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala Gly Leu Ala  
 219 180 185 190  
 221 Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile Val Met Leu  
 222 195 200 205  
 224 Leu Phe Ala Leu  
 225 210

228 (2) INFORMATION FOR SEQ ID NO: 5:

230 (i) SEQUENCE CHARACTERISTICS:

231 (A) LENGTH: 795 base pairs

232 (B) TYPE: nucleic acid

233 (C) STRANDEDNESS: single

234 (D) TOPOLOGY: linear

236 (ii) MOLECULE TYPE: cDNA

238 (iii) HYPOTHETICAL: NO

C--&gt; 240 (iv) ANTI-SENSE: NO

243 (ix) FEATURE:

244 (A) NAME/KEY: CDS

245 (B) LOCATION: 1..792

247 (ix) FEATURE:

248 (A) NAME/KEY: mat\_peptide

249 (B) LOCATION: 1..789

252 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

254 ATG TTG GGT AAG GTC ATC GAT ACC CTT ACA TGC GGC TTC GCC GAC CTC 48  
 255 Met Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu  
 256 1 5 10 15  
 258 GTG GGG TAC ATT CCG CTC GTC GGC GCC CCC CTA GGG GGC GCT GCC AGG 96  
 259 Val Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg  
 260 20 25 30  
 262 GCC CTG GCG CAT GGC GTC CGG GTT CTG GAG GAC GGC GTG AAC TAT GCA 144  
 263 Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala  
 264 35 40 45  
 266 ACA GGG AAT TTG CCC GGT TGC TCT TTC TCT ATC TTC CTC TTG GCT TTG 192  
 267 Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu  
 268 50 55 60  
 270 CTG TCC TGT CTG ACC GTT CCA GCT TCC GCT TAT GAA GTG CGC AAC GTG 240  
 271 Leu Ser Cys Leu Thr Val Pro Ala Ser Ala Tyr Glu Val Arg Asn Val  
 272 65 70 75 80  
 274 TCC GGG ATG TAC CAT GTC ACG AAC GAC TGC TCC AAC TCA AGC ATT GTG 288  
 275 Ser Gly Met Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val  
 276 85 90 95  
 278 TAT GAG GCA GCG GAC ATG ATC ATG CAC ACC CCC GGG TGC GTG CCC TGC 336  
 279 Tyr Glu Ala Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys  
 280 100 105 110  
 282 GTT CGG GAG AAC AAC TCT TCC CGC TGC TGG GTA GCG CTC ACC CCC ACG 384  
 283 Val Arg Glu Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr  
 284 115 120 125  
 286 CTC GCA GCT AGG AAC GCC AGC GTC CCC ACC ACG ACA ATA CGA CGC CAC 432  
 287 Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His  
 288 130 135 140

## RAW SEQUENCE LISTING

DATE: 12/04/2001

PATENT APPLICATION: US/09/973,025

TIME: 11:35:30

Input Set : N:\Crf3\RULE60\09973025.txt

Output Set: N:\CRF3\12042001\I973025.raw

```

290 GTC GAT TTG CTC GTT GGG GCG GCT GCT TTC TGT TCC GCT ATG TAC GTG      480
291 Val Asp Leu Leu Val Gly Ala Ala Ala Phe Cys Ser Ala Met Tyr Val
292 145      150      155      160
294 GGG GAC CTC TGC GGA TCT GTC TTC CTC GTC TCC CAG CTG TTC ACC ATC      528
295 Gly Asp Leu Cys Gly Ser Val Phe Leu Val Ser Gln Leu Phe Thr Ile
296      165      170      175
298 TCG CCT CGC CGG CAT GAG ACG GTG CAG GAC TGC AAT TGC TCA ATC TAT      576
299 Ser Pro Arg Arg His Glu Thr Val Gln Asp Cys Asn Cys Ser Ile Tyr
300      180      185      190
302 CCC GGC CAC ATA ACG GGT CAC CGT ATG GCT TGG GAT ATG ATG ATG AAC      624
303 Pro Gly His Ile Thr Gly His Arg Met Ala Trp Asp Met Met Met Asn
304      195      200      205
306 TGG TCG CCT ACA ACG GCC CTG GTG GTA TCG CAG CTG CTC CGG ATC CCA      672
307 Trp Ser Pro Thr Thr Ala Leu Val Val Ser Gln Leu Leu Arg Ile Pro
308      210      215      220
310 CAA GCT GTC GTG GAC ATG GTG GCG GGG GCC CAT TGG GGA GTC CTG GCG      720
311 Gln Ala Val Val Asp Met Val Ala Gly Ala His Trp Gly Val Leu Ala
312 225      230      235      240
314 GGT CTC GCC TAC TAT TCC ATG GTG GGG AAC TGG GCT AAG GTT TTG ATT      768
315 Gly Leu Ala Tyr Tyr Ser Met Val Gly Asn Trp Ala Lys Val Leu Ile
316      245      250      255
318 GTG ATG CTA CTC TTT GCT CCC TAATAG      795
319 Val Met Leu Leu Phe Ala Pro
320      260
323 (2) INFORMATION FOR SEQ ID NO: 6:
325 (i) SEQUENCE CHARACTERISTICS:
326 (A) LENGTH: 263 amino acids
327 (B) TYPE: amino acid
328 (D) TOPOLOGY: linear
330 (ii) MOLECULE TYPE: protein
332 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
334 Met Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp Leu
335 1      5      10      15
337 Val Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala Arg
338      20      25      30
340 Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr Ala
341      35      40      45
343 Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala Leu
344      50      55      60
346 Leu Ser Cys Leu Thr Val Pro Ala Ser Ala Tyr Glu Val Arg Asn Val
347 65      70      75      80
349 Ser Gly Met Tyr His Val Thr Asn Asp Cys Ser Asn Ser Ser Ile Val
350      85      90      95
352 Tyr Glu Ala Ala Asp Met Ile Met His Thr Pro Gly Cys Val Pro Cys
353      100      105      110
355 Val Arg Glu Asn Asn Ser Ser Arg Cys Trp Val Ala Leu Thr Pro Thr
356      115      120      125
358 Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His
359      130      135      140

```

## VERIFICATION SUMMARY

PATENT APPLICATION: US/09/973,025

DATE: 12/04/2001

TIME: 11:35:31

Input Set : N:\Crf3\RULE60\09973025.txt

Output Set: N:\CRF3\12042001\I973025.raw

L:30 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]  
L:31 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]  
L:60 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:81 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:103 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:240 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:398 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:466 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7  
L:535 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:586 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:9  
L:643 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:752 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:890 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:910 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:930 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:950 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:970 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:990 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1010 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1082 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:21  
L:1154 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1277 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1407 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1544 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1680 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1817 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1837 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:1857 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:2112 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:2304 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:2497 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:2675 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:2855 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:3097 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:3443 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:3659 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:49  
L:4723 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4743 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4765 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4786 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4806 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4826 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4846 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4866 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4886 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4906 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4926 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4946 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]

## VERIFICATION SUMMARY

DATE: 12/04/2001

PATENT APPLICATION: US/09/973,025

TIME: 11:35:31

Input Set : N:\Crf3\RULE60\09973025.txt

Output Set: N:\CRF3\12042001\I973025.raw

L:4966 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:4986 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:5006 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:5026 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:5046 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]  
L:5066 M:220 C: Keyword misspelled or invalid format, [(iv) ANTI-SENSE:]